



Catalogue no. 75F0002MIE — No. 002

ISSN: 1707-2840

ISBN: 0-662-40003-8

Research Paper

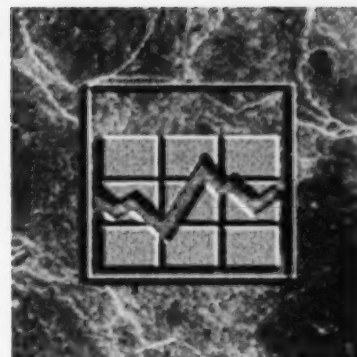
Income Research Paper Series

Recent changes in geography content in the Survey of Labour and Income Dynamics (SLID)

by Chris Li, Gaétan Garneau and Heather Lathe

Income Statistics Division
Jean Talon Building, Ottawa, K1A 0T6

Telephone: 613 951-7355



Statistics
Canada

Statistique
Canada

Canada

How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to: Income Statistics Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 (telephone: (613) 951-7355; (888) 297-7355; income@statcan.ca).

For information on the wide range of data available from Statistics Canada, you can contact us by calling one of our toll-free numbers. You can also contact us by e-mail or by visiting our website.

National inquiries line	1 800 263-1136
National telecommunications device for the hearing impaired	1 800 363-7629
Depository Services Program inquiries	1 800 700-1033
Fax line for Depository Services Program	1 800 889-9734
E-mail inquiries	infostats@statcan.ca
Website	www.statcan.ca

Information to access the product

This product, catalogue no. 75F0002MIE, is available for free. To obtain a single issue, visit our website at www.statcan.ca and select Our Products and Services.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the Agency has developed standards of service that its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll free at 1 800 263-1136.



Statistics Canada
Income Statistics Division

Income research paper series

**Recent changes in geography content in the
Survey of Labour and Income Dynamics (SLID)**

Published by authority of the Minister responsible for Statistics Canada

© Minister of Industry, 2005

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without prior written permission from Licence Services, Marketing Division, Statistics Canada, Ottawa, Ontario, Canada K1A 0T6.

March 2005

Catalogue no. 75F0002MIE2005002

ISSN: 1707-2840

ISBN: 0-662-40003-8

Frequency: Irregular

Ottawa

La version française de cette publication est disponible sur demande (n° 75F0002MIF au catalogue).

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

Abstract

This paper describes the changes made to the structure of geography information on SLID from reference year 1999 onwards. It goes into reasons for changing to the 2001 Census-based geography, shows how the overlap between the 1991 and 2001 Census-based concepts are handled, provides detail on how the geographic concepts are implemented, discusses a new imputation procedure and finishes with an illustration of the impact of these changes on selected tables.

Table of Contents

Introduction.....	6
The postal code: the starting point for geography information.....	7
Reasons for changing to 2001 Census-based design	7
Overlap of the Census-based geography designs in SLID data.....	9
New and discontinued standard geographic concepts	9
Blending of two geography designs in a single set of geography variables	11
How does SLID do geography processing?.....	12
Imputation methods	15
1. Longitudinal imputation – backward	16
2. Longitudinal imputation – forward	16
3. Imputation using six-digit telephone number	16
4. Distance imputation using previous year's postal code.....	17
5. Random imputation of a postal code/DA/CD	17
Impact on data due to the switch to 2001 Census-based design.....	19
Conclusion	21
References.....	22



Introduction

Like many other household surveys, the Survey of Labour and Income Dynamics (SLID) asks its respondents to provide their current address every time they are interviewed. Address information is used not only for survey collection purposes, but also for assigning geographic information during data processing. Detailed geographic information permits data analysis for different types of regions within a province or across provinces.

The norms for coding geographic information in Statistics Canada are set by the Geography Division, primarily for the Census of Population, but also for the benefit of surveys. In preparation for the Census every five years, the Geography Division updates the geography-related lists of Canadian communities, to reflect any changes in their names or administrative boundaries that occurred since the previous Census. The new Census population counts are then used to categorize different areas according to their population density and other geographic characteristics.

Therefore, the SLID data always follow the geographic design of a particular Census. When the survey began collection in January 1993, the 1991 Census-based design was chosen, as it was the most recent one available. There was no update of SLID to the 1996 Census, because the need to change was not yet great enough. But starting with the publication of data for 2002, SLID changed its data processing to use the 2001 Census-based design. Both the cumulative geography changes since 1991 and the introduction of new concepts and methods by Geography Division in the 2001 Census design made it necessary to change the survey processing.

In making the decision to switch to the 2001 Census-based design starting with 2002, the SLID staff decided to ensure that not only the data for 2002 could be analysed using the new geographic design, but also the preceding three years, to provide a consistent four-year time series for the newer geographic detail. This period was a compromise between either revising none of the past years or revising all of them.

While the 1991 Census-based geography variables were maintained on the database, the new 2001 Census-based geography variables were added, with a one-to-one correspondence with the existing variables. It was also decided to maintain identical code sets for both sets of variables. To do this, the SLID staff modified and expanded the code sets to include codes for any new or re-named areas found in the new design, as compared to the 1991 design.

As for the published tables which include geographic detail below the provincial level (usually Census Metropolitan Areas and Census Agglomerations), they were updated so that the years 1999 to 2002 reflect the 2001 design. The estimates for years prior to 1999 were not revised and still reflect the 1991 Census-based design.

Finally, the SLID staff also took advantage of the change from the 1991 to the 2001 Census-based design to introduce improved methods for filling in missing values of

geographic information, caused by missing address information. This process is called imputation. The improved imputation methods were also applied when the data for 1999, 2000 and 2001 were revised.

The postal code: the starting point for geography information

The SLID database contains several types of geographic detail. Although they are not all listed here (they are listed in Table 3: Geography variables on SLID database), some of the most important ones are: the Federal Electoral District (FED), dissemination area (DA), census metropolitan area/census agglomeration (CMA/CA), census division (CD), census subdivision (CSD), economic region (ER), latitude and longitude of residence, and employment insurance region (EIR).¹

The building block for all these types of areas is the postal code. The area represented by a postal code is usually small enough, at least in more urbanized areas, to determine unambiguously all other geographical areas of interest. The postal code is obtained as part of the respondent's address, each time they are interviewed. Fortunately, most respondents know their postal code. Nevertheless, for some households in the sample, it is "missing" when it comes back from collection, due to non-response or a data capture error. For these records, the postal code is derived via the telephone number or it is imputed during survey processing in order to have a complete file. Only then are all the rest of the geographical variables derived.²

Reasons for changing to 2001 Census-based design

Certain geographic classifications need to be updated with every Census. First of all, the categorization of communities into Census Metropolitan Areas (CMAs), Census Agglomerations (CAs) or non-CMA/CA areas depends on their population as determined by the Census of Population, which is conducted every five years. Secondly, Geography Division changes the boundaries from one census to another to include new neighbourhoods or exclude others, the corresponding population changes, and this alone may change the profile of people living there. If the data are not updated to reflect the boundary changes, then they may not be adequately representative of the new area.

Up until reference year 2001, the geographic information on SLID was based on the 1991 Census-based design. Starting with the release of data for reference year 2002, the geographical information of SLID is based on the 2001 Census-based design.

1. For more information on these concepts, please refer to: Census Operations Division, 2002, "2001 Census Dictionary, Reference", Statistics Canada, Catalogue no. 92-378-XPE.

2. There is one drawback associated with the postal code as the sole collection input for geographic coding: it does not always represent the residence of the household. This is particularly the case in rural areas, where rural route service and post office pick-up is commonly used to deliver mail. But, in the majority of cases, the postal code corresponds to the location of the respondent's residence.

There were a few reasons why SLID did not change to the 1996 Census-based design when it became available. Several tests performed on the SLID data showed that the changes from the 1991 to the 1996 Census-based design were not important enough to warrant a change in the SLID geography coding at that time. However, when the 2001 Census-based design became available, it became clear that a change was necessary. Not only were there cumulative changes in population counts over a ten-year period, but also there were many boundary changes throughout the second half of the decade (1996 to 2001)—somewhat more than in the first part of the decade.

To illustrate this second effect, Table 1 below shows the impact of administrative boundary changes on the population of fifteen major CMAs, for the years 1991 to 1996, 1996 to 2001, and the combined ten-year period from 1991 to 2001.

Between 1991 and 1996, there were relatively small changes in population due to boundary adjustments. Out of the 15 CMAs published in SLID's standard tables, only three of them had boundary changes that caused the population estimate to change by one percent or more. Between 1996 and 2001, only four had such significant boundary changes. However, six of the fifteen CMAs were affected between 1991 and 2001, suggesting a change was necessary.

Table 1: Percentage change in population due to boundary changes of CMAs

CMA Name	Population change due to boundary changes between Censuses (%)		
	1991 to 1996	1996 to 2001	1991 to 2001
Montreal	2.7%	0.0%	2.8%
Ottawa-Gatineau	2.4%	-1.2%	1.3%
Winnipeg	1.3%	0.0%	1.3%
Halifax	0.0%	3.3%	3.5%
Québec City	0.0%	0.0%	0.0%
Toronto	0.2%	0.0%	0.2%
Hamilton	0.0%	0.0%	0.0%
St. Catharines-Niagara	0.0%	0.0%	0.0%
Kitchener	0.0%	0.0%	0.0%
London	0.0%	4.7%	4.8%
Windsor	0.0%	3.1%	2.9%
Calgary	0.0%	0.0%	0.0%
Edmonton	0.1%	0.0%	0.2%
Vancouver	0.0%	0.0%	0.0%
Victoria	0.0%	0.0%	0.0%

Notes:

The percentages in the table may not sum up due to rounding.

The 0.0% values in the table implies that there was either no boundary change between the Censuses or the population did not change significantly even though there was a boundary change.

Source: GeoRef 91 (Catalogue no. 92-345D); GeoSuite 2001 (Catalogue no. 92F0085XCB); and Geography Division.

Overlap of the Census-based geography designs in SLID data

In making the switch to using the 2001 Census-based design, the SLID staff decided to historically revise the 1999, 2000 and 2001 sub-provincial data to follow the geographic boundaries of the 2001 Census-based design, while also keeping the earlier values in a separate set of variables. It gives an overlap of three years that can be studied using either the 1991 or 2001 Census geographic boundaries. This choice makes data analysis possible using a uniform geography design for the full six years of each panel of respondents. Table 2 shows how the database is structured.

Table 2: Structure of the geography variables

Reference years:	1992 to 1998	1999 to 2001	2002 on
Variables without 'X' e.g. CMACA25	Refers to 1991 boundaries	Refers to 2001 boundaries	
Variables with 'X' e.g. XCMACA25	Refers to 1991 boundaries		Coded to 'Not Applicable'

All the 1991 Census-based design geography values, covering years 1992 to 2001, are stored in variables that have the same names except they begin with an 'X'. For example, data for CMACA25 was copied to XCMACA25. The regular geography variables contain the 2001 Census-based design values, for the period 1999 onwards.

Most of the geography variables refer to the household's location at the end of the reference year. But there are four variables on geographic mobility that are directly affected by the change to the 2001 Census-based design: SAMCMA26 – for those living in a CMA/CA at then end of a reference year, this is a flag indicating whether the person is living in the same CMA/CA as at the end of the previous year; SAMCSD26 – flag to indicate whether the person is living in the same CSD as at the end of the previous year; DISTM26 – distance moved expressed in kilometres; and MOVST29 – move status for the person for the reference year.

These four variables are derived from the other geography variables by comparing the previous and current reference years' values. Since two years of data are required to derive a value for a given year, they now use the 2001 Census-based design starting with 2000 (which is the comparison of 1999 and 2000). For reference years up to and including 1999, they use the 1991 Census-based design.

New and discontinued standard geographic concepts

The 2001 Census-based design includes three new standard geographic concepts that were not present in either the 1991 or 1996 Census-based designs (or earlier ones).

1. The **dissemination area (DA)**³ is a small, relatively stable geographic unit composed of one or more blocks. Starting with 2001, the dissemination area is the smallest standard geographic area for which any census data can be disseminated (i.e., tabulated for analysis), hence the name chosen for it. Previously, the smallest area for disseminating data was the **enumeration area (EA)**⁴. The EA is a geographic area canvassed by one census representative, and it is still used by the Census for collection purposes.
2. The **census metropolitan area and census agglomeration influenced zone (MIZ)**⁵ is a new concept introduced starting with the 2001 Census-based design. It classifies Census sub-divisions (CSDs) that are not part of either a CMA or CA according to the degree of influence that CMAs and/or CAs collectively have on them. The degree of influence on the CSD is estimated by the extent of the commuting flow between it and other areas—specifically, the proportion of the total employed labour force living in the CSD that works in any CMA/CA urban core. The CSDs with a commuting flow of 30% or more are classified as the “strong” MIZ. The CSDs with a commuting flow of at least 5% and less than 30% are called the “moderate” MIZ. The CSDs with a commuting flow of more than 0%, but less than 5%, are included in the “weak” MIZ category. Those CSDs with fewer than 40 people in the resident labour force commuting to work in CMA/CA urban cores are classified as “no” MIZ.
3. The **Statistical Area Classification (SAC)**⁶ groups census subdivisions (CSDs) according to whether they are a component of a CMA, CA, MIZ or the territories. This more detailed classification replaces the **CMA/CA type**. CMAs and CAs are no longer classified as either “consolidated” or “primary”. Also, CAs (but not CMAs) are classified as “tracted” or “non-tracted”. Finally, all other areas, which did not have any classification before, are now classified using the MIZ.

Table 3 contains the geography variables on the SLID database. The first two columns list the geography variables available on SLID with reference to the 1991 Census-based design. The third column indicates whether the same 1991 Census-based design variables are also available in the 2001 Census-based design.

3. For more information on these concepts, please refer to: Census Operations Division, 2002, “2001 Census Dictionary, Reference”, Statistics Canada, Catalogue no. 92-378-XPE.

4. For more information on these concepts, please refer to: Census Operations Division, 1992, “1991 Census Dictionary, Reference”, Statistics Canada, Catalogue no. 92-301E.

5. For more information on these concepts, please refer to: Census Operations Division, 2002, “2001 Census Dictionary, Reference”, Statistics Canada, Catalogue no. 92-378-XPE.

6. For more information on these concepts, please refer to: Census Operations Division, 2002, “2001 Census Dictionary, Reference”, Statistics Canada, Catalogue no. 92-378-XPE.

Table 3: Geography variables on SLID database

1991 Variable	Description	In 2001 Census-based design
PVRES25	Province of residence	Yes
PVREG25	Province of residence group	Yes
FEDRES25	Federal electoral district ⁷	Yes
EARES25	Enumeration area	No, replaced by dissemination area: DARES25
CDRES25	Census division	Yes
CSDRES25	Census subdivision	Yes
REGRES25	Region	Yes
ERRES25	Economic region	Yes
CMACA25	CMA/CA	Yes
CMA1G25	CMA/CA group 1	Yes
CMA2G25	CMA/CA group 2	Yes
CMAPO25	CMA/CA population	Yes
CATYP25	CMA/CA type	No, replaced by standard area classification type: SACTYP25 (includes MIZ classifications)
LNGRES25	Longitude of residence	Yes
LATRES25	Latitude of residence	Yes
URBRUR25	Urban/rural household	Yes
URBCD25	Urban area code	Yes
URBPO25	Population of area of residence	Yes
URBSZ25	Urban size	Yes
URBSZG25	Size of area of residence	Yes
USZGA25	Adjusted size of area of residence	Yes
EIR25	Employment Insurance Region ⁸	Yes
POSTCD25	Postal code	Yes
POSTCI25	Postal code imputation flag	Yes

Blending of two geography designs in a single set of geography variables

Those geography concepts which are common to both the 1991 and 2001 Censuses continue to exist under the same variable names in SLID as before the 2002 data release. Instead of creating new variables, the SLID staff augmented the categories and descriptions as necessary to incorporate new areas and names.

Any geography categories that only apply to certain years (because they only apply in one of the two Census geography designs) will still appear in the code set, but with a note saying for what years it is not applicable. There will be no households with those codes in those years.

7. FEDs are represented by a member of the House of Commons.

8. For the 1991 Census-based design, the July 1996 delineation of employment insurance regions is used. For the 2001 Census-based design, the July 2000 delineation of employment insurance regions is used.

Below are three examples that show how the variable CMACA25 has been modified to account for the two geography designs. Note that, in order to qualify as a Census Metropolitan Area (CMA) or Census Agglomeration (CA), a municipality must satisfy certain population criteria; CAs can be created or retired between Census designs, and although CMAs are never retired, there could be new ones created. Households in non-CMA/CA areas are given a value of "not applicable" for variable CMACA25.

Case A: Codes only applicable prior to 1999 (the 2001 Census-based design). The CMACA25 code 730 represents Weyburn, Saskatchewan. It was a CA in the 1991 Census-based design, but was retired in 1996 because the population fell below 10,000 according to the Census results (it retained its non-CMA/CA status in 2001). Households in Weyburn are no longer part of a CMA/CA, so they would be given the "not applicable" code for this variable starting in 1999. An example of a CMA/CA that disappeared because of a boundary change is Saint-Jérôme, Quebec, code 475. It was a CA on its own in the 1991 Census-based design, but by the time of the 2001 Census it had become amalgamated with Montreal, which is already its own CMA. Unlike Weyburn, the households in Saint-Jérôme are still part of a CMA/CA and would be given the code for Montréal starting in 1999.

Case B: Codes only applicable for years 1999 and on (the 2001 Census-based design). An example of a CA that qualified in the 2001 Census-based design but not in the 1991 design is Squamish, British Columbia. A new code was created for it, code 934.

Case C: Codes applicable in all reference years. Most codes have retained the same name and still apply because that area is still a CMA or CA on its own. The boundary may also have changed if there was a change in municipal jurisdictions; this is not indicated in the CMACA25 variable, but it is taken into account in the Census-specific mapping of CSDs to CMAs, CAs, and non-CMA-CA areas. The CMACA25 code 225 used to be named Sydney, Nova Scotia in the 1991 Census-based design, but it was renamed to Cape Breton due to amalgamation and municipal restructuring. This is an example of a name change combined with a boundary change. The description with this code indicates that it represented Sydney until 1998, and thereafter Cape Breton.

How does SLID do geography processing?

This section describes the input files and processing steps used by SLID for 2002 and also retroactively for 1999, 2000, and 2001 in the revised data.⁹

9. For information on how SLID processed its geographical information using the 1991 Census-based design, please refer to "SLID Geography and its Impact on Low Income Measurement" by Cunningham, R., Lafrance, P., Rowland, J. and Murray, J., 1991. The Income and Labour Dynamics Working Paper Series, Statistics Canada, Catalogue no. 97-09.

Input files

There are four main input files required to process geographical information for each household. They are the Postal Code Conversion File (PCCF), the Geographic Attribute File (GAF), the Employment Insurance Region (EIR) file and the Telephone Billing file.

The **Postal Code Conversion File (PCCF)** produced by Geography Division contains the link between the six-character postal codes and Statistics Canada's standard census geographical areas (such as DAs, municipalities and CMA/CAs). It also provides the latitude and longitude coordinates for a point representing the approximate location of the center of the postal code area. In addition to being updated every five years to reflect new Census geography, the PCCF is updated every six months to take into account the postal code changes continually introduced by Canada Post Corporation (CPC).¹⁰

The **Geographic Attribute File (GAF)** contains the population and dwelling counts of the latest census, by Dissemination Area (DA). It also provides the correspondence or "link" between each DA and all higher geographic levels except the Employment Insurance Region. This includes information identifying urban and rural areas (urban/rural flag, urban area code and urban area population) and the longitude and latitude of a representative point in each DA.

The **Employment Insurance Region (EIR)** file provides the link between each DA and Employment Insurance Regions (EIRs), which are defined by Human Resources and Skills Development Canada (HRSD) and Social Development Canada (SD) (formerly Human Resources Development Canada). Eligibility for Employment Insurance benefits is based, in part, on the unemployment rate in the applicant's EIR.

The above three files (PCCF, GAF, and EIR) can be merged as one file, using the dissemination area as their common field. Whether the SLID data are linked to these files one at a time or after merging the three together does not really matter, for the purposes of simplifying the description of processing steps below, we can refer to them jointly as the PCCF-GAF-EIR file.

The SLID staff produces a single **Telephone Billing file** each year from the multiple telephone billing files received quarterly from approximately 22 companies that provide telephone services across Canada. Note that these telephone billing files do not contain the names of the subscribers nor any billing information. Only the telephone number, postal code and province fields are required for SLID processing.¹¹

10. The PCCF was first linked to the 1981 Census geographic areas and has undergone four "conversions" since then, to reflect the 1986, 1991, 1996 and 2001 censuses. For more information, please refer to "Postal Code Conversion File (PCCF), Reference Guide", Statistics Canada, Catalogue no. 95F0153GIE.

11. Although useful, the telephone files have a few limitations which could prevent a link with survey respondents, or which might give an address other than the person's permanent residence. The bill may be sent to another residence or a post office, although this does not actually matter unless the postal code of that address is not the same as the person's residence. Depending on the telephone company, unlisted numbers may not be included, in which case a match would not be found.

Once these file have been obtained, the SLID geography processing for each household can proceed.

Steps to assign geography values

Step 1

After setting aside all records with no postal code in the current year (these will be processed in step 3 later), a match is done between households in the current year and households of the previous year by household ID, province and postal code. For all current-year households for which a match is found, census geographic codes are copied over from the previous year.

Before the unmatched records from this step are passed to the next step, they are checked for consistency of their postal code and telephone number with their province code. For example, households residing in Ontario must have a postal code that begins with a K, L, M, N or P, and their telephone area code must begin with 289, 416, 519, 613, 647, 705, 807, or 905. If either the postal code or the telephone number fails this check, then both are set to "missing" and only the province is retained. Either their geography variables will be assigned values in Step 3, or they will be imputed.

Step 2

Match the unmatched records from Step 1 with the PCCF-GAF-EIR file by province and postal code. All records with a postal code will find a match on the file, unless some invalid codes were not caught in Step 1. For simplicity, we can assume that all unmatched records at this stage have a postal code value of "missing".

Step 3

Match the unmatched records from Step 2 and the records with no postal code (they were set aside in Step 1) with the Telephone Billing file, by their 10-digit telephone number. For successfully matched records, the postal code is added and then the record is re-matched with the PCCF-GAF-EIR file to obtain all the geographic values. Those that did not match at this stage are sent to Step 4.

Step 4

Those records that still do not have a postal code after Step 3 are compared to last year's file at the same stage (i.e. just after Step 3) by household ID, address *excluding* postal code and phone number. If they are the same, then the geographic values are carried forward to the current year, to avoid changing the location of residence through re-imputation. The rest of the records will have their postal code value assigned by imputation.

Step 5

Impute the remaining unmatched records from Step 4. Please refer to the next section "Imputation methods". Once a postal code, CD and DA codes have been imputed, we match these records to the PCCF-GAF-EIR file to retrieve the rest of the geography values.

Geography processing is then complete except for some variables that are not directly provided by the PCCF-GAF-EIR file but which can be derived, such as the distance moved.

Table 4 gives the record counts for the processing steps above.

Table 4: Records assigned geography codes, by method (2001 Census-based design)

Reference Year	1999	2000	2001	2002
Total number of households in sample	35,761 (100%)	37,125 (100%)	36,444 (100%)	36,766 (100%)
Outside Canada records: All fields set to not applicable	305 (1%)	389 (1%)	481 (1%)	307 (1%)
Step 1: Codes copied from previous year, since the postal code and household ID are the same	*	25,491 (69%)	27,572 (76%)	14,646 (40%)
Step 2: Codes obtained by matching postal code to PCCF-GAF-EIR file, since postal code is already available and valid	34,064 (95%)	3,585 (10%)	6,084 (17%)	19,875 (54%)
Step 3: Codes obtained by using telephone number to get a postal code, then matched to PCCF-GAF-EIR file	800 (2%)	4,916 (13%)	728 (2%)	860 (2%)
Step 4: Codes copied from previous year's values since postal code missing but the address and household ID are the same	*	81 (0%)	180 (0%)	247 (1%)
Step 5: Postal code, CD and DA are imputed and then record is matched to PCCF-GAF-EIR file	592 (2%)	2,663 (7%)	1,399 (4%)	831 (2%)

*Not applicable because 1999 was the earliest year to be processed using the 2001 Census-based geography design.

Note: In step 1, the percentage of records having their geography codes copied from the previous year for 2002 is lower than for 2000 and 2001 because of the introduction of a new panel in 2002.

Imputation methods

The postal code imputation methods are presented below in the same order as they are used in SLID processing. These methods were used each year starting with processing of data for 2002, but they also apply to years 1999, 2000 and 2001 because the imputed geography data of those years were re-imputed in a historical revision. Prior to 2002 processing, SLID used the very simple method of randomly selecting any one of the postal codes that existed in the province.

The imputation methods described here are all within a given province. One main reason to restricting imputation to be within province was to respect the weighting procedure for the historical revisions of 1999 to 2001. Another reason for restricting imputation to be within a given province is because some records need their province to be imputed beforehand as the SLID staff was unable to trace them. The method used for province imputation is just to assume that the household still lives in the same province as the last known province.¹²

1. Longitudinal imputation – backward (POSTCI25=40)

- This type of imputation can only be used when doing historical revisions since it takes known information from one year and applies it to one or more preceding years.
- This method was used for the historically revised years, 1999 to 2001.
- If a household record exists in the survey in at least two consecutive years and their address in the second year is known, but their address in the first of the two years is unknown, then it is assumed that the SLID staff lost contact with the respondent in the first year of missing data. This assumption is logical because often when the SLID staff cannot trace a respondent, it is because they moved. Another assumption is applied that they did not move between the two years, so the later year's address information is imputed to the previous year.¹³

2. Longitudinal imputation – forward (POSTCI25=43)

- Only used for reference years 2000 and on – after re-processing 1999 data.
- If the household needs their postal code to be imputed two years consecutively, then they are given the same postal code for these two years. This is because no information is known whether the household did or did not move. The SLID staff did not want to re-impute a postal code for them and re-locate the household. This necessarily assumes that they did not move.

3. Imputation using six-digit telephone number (area code + exchange bank) (POSTCI25=41)

- For those with a telephone number in the current reference year, extract the first six digits (area code + exchange bank). For example, the first six digits of a telephone number for someone residing in Ottawa might be (613744), where (613) is the area code for Ottawa and (744) is the exchange bank. We match these with the six-digit telephone numbers on the telephone billing file which,

12. The imputation of the household location within province of the last known address is an artificial limitation since in reality it is possible that respondents move to another province. Ideally, inter-provincial movement would be allowed for in the imputation method. But since it requires a lot of research into different methods before it can be implemented, it was excluded from the recent changes in geography processing, which already included a new Census design, improved imputation methods, and historical revisions. However, the intention is to look at it in the near future.

13. Postal codes were not affected by the switch in geography design.

was first merged with the PCCF. If there are many postal codes for that area code and exchange bank, we randomly choose one of them.

- If in the current reference year the telephone number is unidentifiable or there is no telephone number for the respondent, then method 4 is used.

4. Distance imputation using previous year's postal code (POSTCI25=42)

- This method is used for records that were in the survey the previous year and did not have to be imputed that year; therefore, one knows with certainty what their location was one year ago. The first stage to this method is to impute how far they have moved. An assumption is made that they had moved because the SLID staff was not able to trace them for the survey. After this first stage, one knows all the possible longitude and latitude coordinates they could have moved to, which would be consistent with the distance moved. The second stage is to select one of these pairs of coordinates and then use it to obtain the postal code. Below are a few more details on how the distance imputation is done.
- For each SLID respondent that was in the survey in both the current and preceding years with a valid postal code in both those years (i.e. without imputation) and whose postal code changed, their "distance moved" is calculated. All these values are sorted by province and move type (i.e. urban-urban, urban-rural, rural-urban or rural-rural). Then, for each household requiring imputation, randomly select one of the values of "distance moved" from all those with the same province and move type as the household. This will be the donor used to establish approximately how far the household requiring imputation has moved.
- The PCCF is used to obtain the list of longitude and latitude coordinates of all the postal code areas within the province; it gives just one pair of coordinates per postal code area. Then, for each household requiring imputation, find all the possible destinations that they could have moved to within the same province given their distance moved and previous location. In order to get a number of destinations, the distance moved is extended to be within, for example, 5 kilometres plus/minus around the donor "distance moved". One of these destinations is randomly chosen to be the destination that the household actually moved to. The postal code can then be obtained from the coordinates.

5. Random imputation of a postal code/DA/CD (POSTCI25=44)

- This last resort method is applied only if no postal code can be found from any of the above methods. A postal code within province is randomly chosen.
- Records that are left for random imputation are usually the ones with no postal code or telephone number. For many records, they only contained their household ID and their province of residence while all the other fields of information are empty.

The variable POSTCI25 serves to record how each household record obtained its final postal code value. The various categories are shown in Table 5 below. The possible categories include the different types of derivation or imputation that were applied,

depending on the year. Codes 30, 31 and 32 now apply only to reference years 1992 to 1998, which are still based on the 1991 Census geography in SLID. Codes 40 to 44 make reference to the 2001 Census geography and SLID's updated imputation methods. Code 40 applies to only 1999 to 2001 and code 43 only applies to 2000 onwards. Also, codes 41, 42 and 44 apply to the years starting with 1999.

Table 5: Categories of variable POSTCI25: How postal code was assigned

Code	Description
10	Collected postal code
20	Postal code derived from telephone number
30	Federal electoral district/Enumeration area assigned by Geography Division using address (RY1992-RY1998)
31	Federal electoral district/Enumeration area imputed by Geography Division using province (RY1992-RY1998)
32	Federal electoral district/Enumeration area assigned from Stratum-Type-Cluster of LFS Frame (RY1992-RY1998)
40	Longitudinal imputation – backward (RY1999-RY2001)
41	Imputation using reference year 6 digit telephone number (area code + exchange bank) (i.e. 613-951) (from RY1999 on)
42	Distance imputation using previous year's postal code (from RY1999 on)
43	Longitudinal imputation – forward (from RY2000 on)
44	Random imputation of postal code/DA/CD (from RY1999 on)

Note: RY stands for reference year. The reference years indicated in brackets implies it is used only for those reference years.

Table 6 below contains the number of records with postal codes imputed by the different methods.

Table 6: Counts by imputation methods (POSTCI25)

Code	Imputation methods	1999	2000	2001	2002
40	Longitudinal imputation – backward (RY1999-RY2001)	41 (6.9%)	1,031 (38.7%)	40 (2.9%)	...
41	Imputation using reference year 6 digit telephone number (i.e. 613-951) (from RY1999 on)	187 (31.6%)	793 (29.8%)	146 (10.4%)	179 (21.5%)
42	Distance imputation using previous year's postal code (from RY1999 on)	303 (51.2%)	751 (28.2%)	982 (70.2%)	483 (58.1%)
43	Longitudinal imputation – forward (from RY2000 on)	...	0 (0.0%)	151 (10.8%)	20 (2.4%)
44	Random imputation of a POSTCD/DA/CD (from RY1999 on)	61 (10.3%)	88 (3.3%)	80 (5.7%)	149 (17.9%)
...	Total	592 (100.0%)	2,663 (100.0%)	1,399 (100.0%)	831 (100.0%)

Note: RY stands for reference year. The reference years indicated in brackets implies it is used only for those reference years.

While introducing these improved methods for replacing missing values of geographic information, the SLID staff noticed an additional improvement that could be made.

The distance imputation method described above for 1999 to 2002 used an approximate formula:

$$(1) DM = 75 * \text{SQRT}[(LATRES_{RY} - LATRES_{RY-1})^2 + (LNGRES_{RY} - LNGRES_{RY-1})^2],$$

where DM=distance moved
 SQRT=square root
 LATRES=latitude of residence
 LNGRES=longitude of residence
 RY=current reference year
 RY-1=previous reference year

This formula (1) tends to underestimate the distance moved by approximately 33% for low latitudes and overestimate the distance moved by approximately 20% for high latitudes.

Starting in 2003, it was replaced by a more precise one:

$$(2) DM = 111.32 * \arccos [\sin(LATRES_{RY}) * \sin(LATRES_{RY-1}) + \cos(LATRES_{RY}) * \cos(LATRES_{RY-1}) * \cos(LNGRES_{RY} - LNGRES_{RY-1})]$$

This formula (2) calculates the distance along a “great circle arc”, i.e., the shortest surface distance between any two points on a sphere (i.e. assuming the earth to be a perfect sphere).

Impact on data due to the switch to 2001 Census-based design

The data prior to 1999 is based on the 1991 Census-based design. Since the data for reference years 1999 and later is based on the 2001 Census-based design, there is a structural break in the data series between 1998 and 1999. Since many geographical boundaries have changed, as well as many name and code changes have occurred, the geography codes for some households between 1998 and 1999 may be different even if the household did not move.

Furthermore, due to the change in imputation methods, some households that were imputed in reference years 1999, 2000 or 2001 may be located in a different location when examining the data for both the 1991 and 2001 Census-based design.

Updating geographical information can impact low income rates in particular. For example, the family size and the community size are used to find the appropriate cutoff. Then the family income is compared to that cutoff. If a family low-income rate is being calculated, then the family is counted as being in low income if its income is less than the cutoff. If a person low-income is being calculated, then all persons in the family are counted as being in low income if the family income is less than the cutoff.

Low income cutoffs (LICOs) are established using data from the Family Expenditure Survey, now known as the Survey of Household Spending. LICOs convey the income level at which a family may be in straitened circumstances because it has to spend a greater proportion of its income on necessities than the average family of similar size. Specifically, the threshold is defined as the income below which a family is likely to spend 20 percentage points more than the average family. There are separate cutoffs for seven sizes of family – from unattached individuals to families of seven or more persons – and for five community sizes – from rural areas to urban areas with a population of more than 500,000.

Table 7 below contains low-income rates (the proportion of persons living in low income) for reference years 1999, 2000 and 2001 for both the 1991 and 2001 Census-based design. This table indicates that the low-income rates are not significantly affected by the change from using the 1991 to the 2001 Census-based design to process the geography variables.

Table 7: Low-income rates for all persons in 15 major CMAs in Canada

	1999			
	Before-tax		After-tax	
	1991 Census	2001 Census	1991 Census	2001 Census
Census-based design				
Montreal	23.9%	24.1%	19.3%	19.4%
Ottawa-Gatineau	17.7%	17.9%	14.6%	14.7%
Winnipeg	19.2%	19.3%	15.7%	15.9%
Halifax	17.1%	17.1%	13.3%	13.5%
Quebec City	18.2%	18.6%	14.0%	14.4%
Toronto	14.8%	14.8%	11.2%	11.2%
Hamilton	12.1%	12.2%	9.2%	9.2%
St. Catherines-Niagara	9.6%	9.5%	5.5%	5.5%
Kitchener	10.8%	10.8%	8.0%	8.0%
London	16.0%	16.1%	12.8%	12.8%
Windsor	12.8%	13.4%	8.1%	8.3%
Calgary	13.5%	13.5%	11.2%	11.3%
Edmonton	15.8%	15.8%	12.5%	12.5%
Vancouver	20.2%	20.2%	16.5%	16.6%
Victoria	15.4%	15.2%	10.6%	10.6%
	2000			
	Before-tax		After-tax	
	1991 Census	2001 Census	1991 Census	2001 Census
Census-based design				
Montreal	21.2%	21.7%	17.6%	17.6%
Ottawa-Gatineau	14.4%	14.8%	12.5%	12.7%
Winnipeg	18.3%	18.8%	14.8%	15.3%
Halifax	14.7%	15.0%	12.0%	12.2%
Quebec City	16.7%	17.3%	13.8%	14.1%
Toronto	13.3%	13.4%	10.5%	10.7%
Hamilton	11.5%	11.5%	9.0%	9.1%
St. Catherines-Niagara	7.5%	7.4%	5.5%	5.5%
Kitchener	10.0%	10.0%	8.1%	8.1%
London	12.0 %	12.0%	9.8%	9.4%
Windsor	11.6%	11.4%	7.8%	7.8%

Calgary	12.2%	12.2%	8.8%	9.0%
Edmonton	17.0%	17.0%	12.2%	12.2%
Vancouver	17.3%	17.3%	14.4%	14.5%
Victoria	18.3%	18.3%	14.8%	13.9%
	2001			
	Before-tax		After-tax	
Census-based design	1991 Census	2001 Census	1991 Census	2001 Census
Montreal	19.0%	19.0%	14.8%	14.9%
Ottawa-Gatineau	11.4%	11.6%	8.1%	8.3%
Winnipeg	16.1%	16.5%	11.2%	11.6%
Halifax	14.2%	14.4%	11.7%	11.7%
Quebec City	17.2%	17.7%	13.1%	13.6%
Toronto	11.0%	11.1%	8.4%	8.4%
Hamilton	10.6%	10.6%	6.9%	6.9%
St. Catharines-Niagara	7.4%	7.4%	4.8%	4.5%
Kitchener	8.1%	8.1%	4.5%	4.5%
London	11.6%	11.6%	7.9%	7.9%
Windsor	9.4%	9.6%	7.9%	7.9%
Calgary	11.7%	11.7%	8.9%	9.0%
Edmonton	15.1%	15.2%	9.8%	9.8%
Vancouver	17.4%	17.4%	13.3%	13.3%
Victoria	14.5%	14.5%	10.1%	10.1%

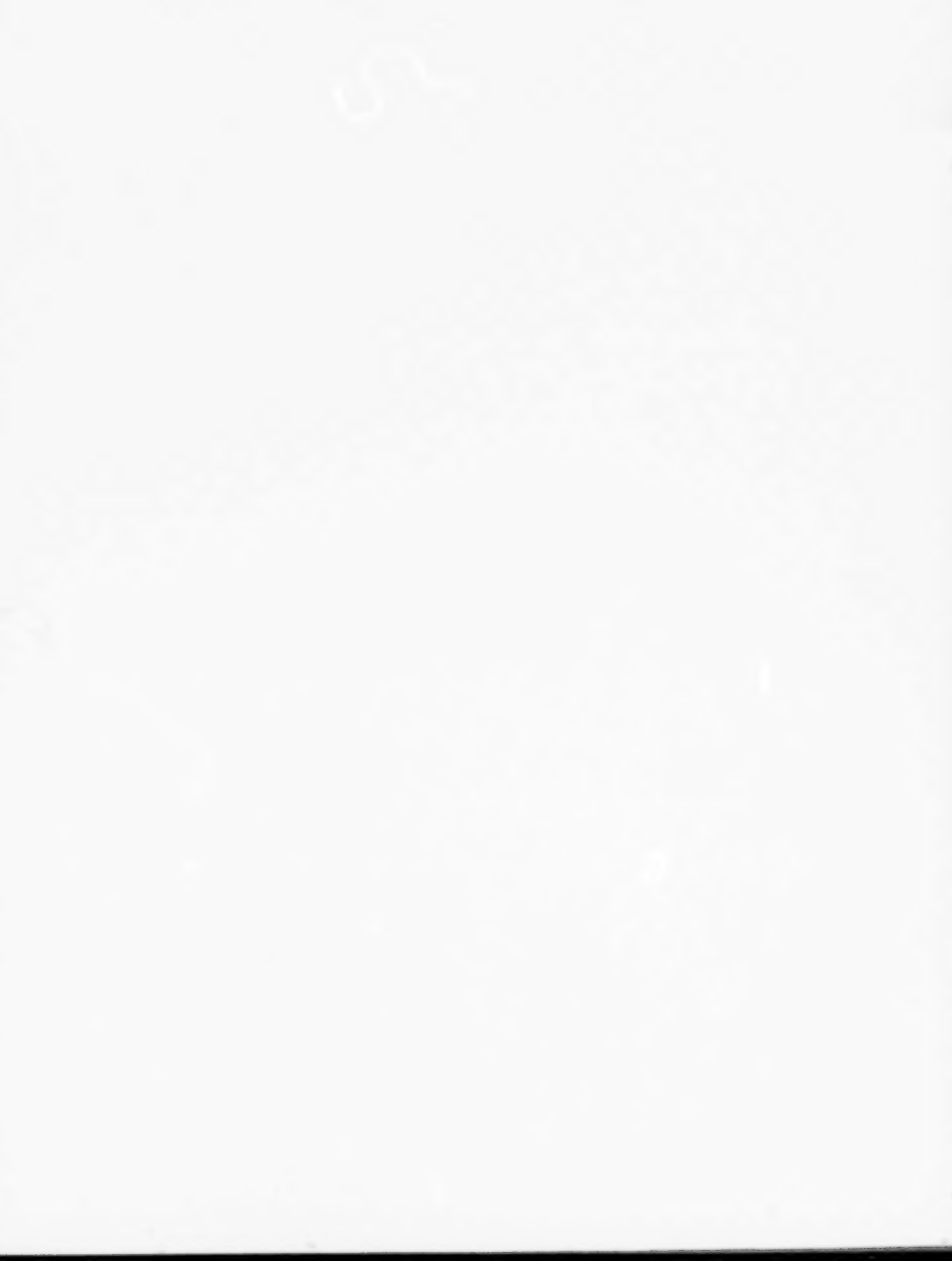
Conclusion

The SLID staff knew at some point that it would be necessary to align the SLID geography definitions to that of a more recent Census. To that end, the SLID geography definitions were changed from 1999 onward, while overlapping the definitions for 1999, 2000 and 2001. The impact, as seen in the previous section, is modest and will allow the SLID staff more breathing room until the next change, perhaps following the 2006 or 2011 Census.

Though imputation was emphasised in this document, the grand majority of geographic coding is done by copying codes from a previous year or by matching a respondent's valid postal code to geography files. The rest are handled by fairly sophisticated methods which ensure that the user receives the best geographic coding that the SLID staff can give them.

References

- Census Operations Division. 1992. "1991 Census Dictionary, Reference". Statistics Canada, Catalogue No. 92-301E.
- Census Operations Division. 1999. "1996 Census Dictionary, Final Edition Reference". Statistics Canada, Catalogue No. 92-351-UPE.
- Census Operations Division. 2002. "2001 Census Dictionary, Reference". Statistics Canada, Catalogue No. 92-378-XPE.
- Cunningham, R., Lafrance, P., Rowland, J. and Murray, J. 1997. "Slid Geography and its Impact on Low Income Measurement." The Income and Labour Dynamics Working Paper Series, Statistics Canada, Catalogue No. 97-09.
- Geography Division. 1995. "Postal Code Conversion File (PCCF), User Guide." Statistics Canada.
- Geography Division. 2003. "Postal Code Conversion File (PCCF), Reference Guide." Statistics Canada, Catalogue No. 92F0153GIE.
- Income Statistics Division. 2002. "Income in Canada, 2002." Statistics Canada, Catalogue No. 75-202-XIE.
- Services, Publications Division. 1992. "Standard Geographical Classification SGC 1991, Volume 1, The Classification." Statistics Canada, Catalogue No. 12-571.
- Standards Division. 2002. "Standard Geographical Classification SGC 2001, The Classification." Statistics Canada, Catalogue No. 12-571-XPB.





N° 75F0002MIF au catalogue — N° 002

ISSN: 1707-2867

ISBN: 0-662-79630-6

Document de recherche

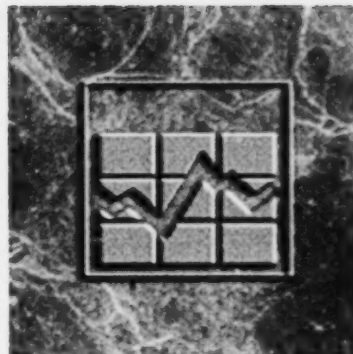
Série de documents de recherche - Revenu

Modifications récemment apportées au volet géographique de l'Enquête sur la dynamique du travail et du revenu (EDTR)

par Chris Li, Gaétan Garneau et Heather Lathe

Division de la statistique du revenu
Immeuble Jean-Talon, Ottawa, K1A 0T6

Téléphone: 1 613 951-7355



Statistique
Canada

Statistics
Canada

Canada

Comment obtenir d'autres renseignements

Toute demande de renseignements au sujet du présent produit ou au sujet de statistiques ou de services connexes doit être adressée à : Division de la statistique du revenu, Statistique Canada, Ottawa, Ontario, K1A 0T6 (téléphone : (613) 951-7355; (888) 297-7355; revenu@statcan.ca).

Pour obtenir des renseignements sur l'ensemble des données de Statistique Canada qui sont disponibles, veuillez composer l'un des numéros sans frais suivants. Vous pouvez également communiquer avec nous par courriel ou visiter notre site Web.

Service national de renseignements	1 800 263-1136
Service national d'appareils de télécommunications pour les malentendants	1 800 363-7629
Renseignements concernant le Programme des bibliothèques de dépôt	1 800 700-1033
Télécopieur pour le Programme des bibliothèques de dépôt	1 800 889-9734
Renseignements par courriel	infostats@statcan.ca
Site Web	www.statcan.ca

Renseignements pour accéder au produit

Le produit n° 75F0002MIF au catalogue est disponible gratuitement. Pour obtenir un exemplaire, il suffit de visiter notre site Web à www.statcan.ca et de choisir la rubrique Nos produits et services.

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois, et ce, dans la langue officielle de leur choix. À cet égard, notre organisme s'est doté de normes de service à la clientèle qui doivent être observées par les employés lorsqu'ils offrent des services à la clientèle. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1 800 263-1136.



Statistique Canada
Division de la statistique du revenu

Série de documents de recherche - Revenu

**Modifications récemment apportées au
volet géographique de l'Enquête sur la
dynamique du travail et du revenu (EDTR)**

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2005

Tous droits réservés. Il est interdit de reproduire ou de transmettre le contenu de la présente publication, sous quelque forme ou par quelque moyen que ce soit, enregistrement sur support magnétique, reproduction électronique, mécanique, photographique, ou autre, ou de l'emmagasiner dans un système de recouvrement, sans l'autorisation écrite préalable des Services de concession des droits de licence, Division du marketing, Statistique Canada, Ottawa, Ontario, Canada K1A 0T6.

Mars 2005

N° 75F0002MIF2005002 au catalogue

ISSN: 1707-2867

ISBN: 0-662-79630-6

Périodicité : Irrégulier

Ottawa

This publication is available in English upon request (Catalogue no. 75F0002MIE)

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population, les entreprises, les administrations canadiennes et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques précises et actuelles.

Résumé

Le présent document décrit les modifications apportées à la structure des données géographiques de l'EDTR depuis l'année de référence 1999. Le document traite des raisons pour lesquelles il y a eu des modifications en fonction des données géographiques du Recensement de 2001, montre comment est traité le chevauchement entre les concepts géographiques des recensements de 1991 et de 2001, explique en détail la manière dont les concepts géographiques sont appliqués, analyse une nouvelle procédure d'imputation, puis illustre en conclusion l'effet de ces modifications sur certains tableaux.

Table des matières

Introduction.....	6
Le code postal : point de départ des données géographiques	7
Raisons du passage aux concepts géographiques du Recensement de 2001	8
Chevauchement des conceptions géographiques du recensement dans les données de l'EDTR.....	9
Concepts géographiques normalisés nouveaux et abandonnés.....	10
Méthodes d'imputation	17
1. Imputation longitudinale – rétrospective	18
2. Imputation longitudinale – prospective	18
3. Imputation à l'aide du numéro de téléphone à six chiffres	18
4. Imputation de la distance à l'aide du code postal de l'année précédente	19
5. Imputation au hasard du code postal/de l'AD/de la DR	20
Passage aux concepts géographiques du Recensement de 2001 : l'effet sur les données	22
Conclusion	24
Ouvrages de référence.....	25

Introduction

À l'instar de nombreuses autres enquêtes-ménages, l'Enquête sur la dynamique du travail et du revenu (EDTR) exige des répondants qu'ils fournissent leur adresse actuelle lors de chaque interview. Les renseignements sur l'adresse servent non seulement à la collecte, mais aussi à l'attribution des renseignements géographiques au cours du traitement des données. Les renseignements géographiques détaillés permettent l'analyse de données des différents types de régions au sein d'une province ou entre provinces.

Les normes de codage des données géographiques à Statistique Canada sont établies par la Division de la géographie, essentiellement aux fins du Recensement de la population, mais également aux fins des enquêtes. En vue du recensement quinquennal, la Division de la géographie met à jour les listes géographiques des collectivités canadiennes, de façon à ce que ces listes reflètent les modifications apportées aux noms ou aux limites administratives depuis le recensement précédent. Les nouveaux chiffres de population du recensement servent alors au classement par catégories des différentes zones selon leur densité démographique et d'autres caractéristiques géographiques.

Par conséquent, les données de l'EDTR sont constamment conformes aux concepts géographiques d'un recensement particulier. Lorsque la collecte des données de l'Enquête a débuté en janvier 1993, on a choisi les concepts géographiques du Recensement de 1991 puisque celui-ci était le plus récent. Il n'y a pas eu de mise à jour de l'EDTR par rapport au Recensement de 1996, parce que le besoin de modifications n'était pas encore assez manifeste. Toutefois, depuis la publication des données de 2002, on a modifié le traitement des données de l'EDTR pour adopter les concepts géographiques du Recensement de 2001. Tant les modifications géographiques cumulatives depuis 1991 que l'application par la Division de la géographie de nouvelles méthodes et de nouveaux concepts géographiques au Recensement de 2001 ont nécessité la modification du traitement des données de l'Enquête.

Dans sa décision de passer aux concepts géographiques du Recensement de 2001 à compter de 2002, l'équipe de l'EDTR a voulu faire en sorte que non seulement les données de 2002, mais aussi celles des trois années précédentes puissent être analysées à l'aide de la nouvelle conception géographique, et ce, afin d'obtenir une série chronologique cohérente de quatre ans pour les détails géographiques les plus récents. Le choix de cette période est un compromis entre l'absence de révision des années antérieures et la révision intégrale de celles-ci.

Tout en gardant les variables géographiques associées au Recensement de 1991 dans la base de données, nous avons ajouté les nouvelles variables géographiques du Recensement de 2001, avec la correspondance univoque aux variables existantes. De plus, on a décidé de maintenir les mêmes ensembles de codes pour les deux séries de variables. Pour ce faire, l'équipe de l'EDTR a modifié et élargi les ensembles de

codes pour y insérer les codes des zones nouvelles ou renommées qui se trouvent dans la nouvelle conception géographique, comparativement à celle de 1991.

Quant aux tableaux publiés comportant les détails géographiques en deçà de l'échelon provincial (habituellement les régions métropolitaines de recensement et les agglomérations de recensement), ils ont été mis à jour de façon à ce que les années 1999 à 2002 reflètent les concepts géographiques du Recensement de 2001. Les estimations des années antérieures à 1999 n'ont pas été révisées et reflètent encore les concepts géographiques du Recensement de 1991.

Enfin, l'équipe de l'EDTR a également profité du passage des concepts géographiques du Recensement de 1991 à ceux de 2001 pour appliquer des méthodes améliorées afin de combler les valeurs manquantes des données géographiques causées par les données manquantes pour ce qui est de l'adresse. Ce processus se nomme imputation. Les méthodes améliorées d'imputation ont aussi été appliquées lors de la révision des données de 1999, 2000 et 2001.

Le code postal : point de départ des données géographiques

La base de données de l'EDTR comporte des détails géographiques de plusieurs types. Bien que ces détails ne soient pas tous énumérés ici (ils le sont au tableau 3 : Variables géographiques dans la base de données de l'EDTR), certains parmi les plus importants sont les suivants : la circonscription électorale fédérale (CÉF), l'aire de diffusion (AD), la région métropolitaine de recensement/l'agglomération de recensement (RMR/AR), la division de recensement (DR), la subdivision de recensement (SDR), la région économique (RÉ), la latitude et la longitude de résidence, de même que la région d'assurance-emploi (RAE)¹.

L'unité élémentaire de tous les types de régions est le code postal. La région circonscrite par le code postal est habituellement de superficie assez restreinte, du moins dans les régions plus urbanisées, afin de déterminer sans ambiguïté tous les autres domaines géographiques d'intérêt. On tire le code postal de l'adresse du répondant, obtenue lors de chaque interview. Heureusement, la plupart des répondants connaissent leur code postal. Néanmoins, le code postal de certains ménages de l'échantillon est « manquant » une fois la collecte terminée, à cause de la non-réponse ou d'une erreur de saisie des données. Aux fins de ces enregistrements, le code postal est dérivé du numéro de téléphone ou imputé lors du traitement de

1. Pour en savoir plus sur ces concepts, reportez-vous au document : Division des opérations du recensement, *Dictionnaire du recensement de 2001, Référence*, n° 92-378-XPB au catalogue de Statistique Canada, 2002.

l'enquête afin d'obtenir un dossier complet. Ce n'est qu'à ce moment que les autres variables géographiques sont dérivées².

Raisons du passage aux concepts géographiques du Recensement de 2001

Lors de chaque recensement, il faut mettre à jour certaines classifications géographiques. D'abord, le classement par catégories des collectivités dans les régions métropolitaines de recensement (RMR), les agglomérations de recensement (AR) ou les régions autres que les RMR/AR dépend de leur population telle que déterminée par le Recensement, lequel a lieu tous les cinq ans. Ensuite, la Division de la géographie modifie les limites d'un recensement à l'autre afin d'englober de nouveaux quartiers ou d'en exclure d'autres, suivant les modifications démographiques correspondantes. Il en résulte la variation possible du profil des personnes habitant tel ou tel quartier. Sans mise à jour permettant de refléter les modifications aux limites, les données risquent de ne pas représenter adéquatement la nouvelle région.

Jusqu'à l'année de référence 2001, les renseignements géographiques figurant dans l'EDTR étaient fondés sur les concepts géographiques du Recensement de 1991. Depuis la publication des données de l'année de référence 2002, les renseignements géographiques de l'EDTR sont fondés sur les concepts géographiques du Recensement de 2001.

Quelques raisons avaient empêché le passage de l'EDTR aux concepts géographiques du Recensement de 1996 quand celui-ci était prêt. Plusieurs tests effectués sur les données de l'EDTR ont montré que les modifications apportées aux concepts géographiques du Recensement de 1996 à partir de celles de 1991 n'étaient pas assez importantes pour justifier une modification du codage géographique de l'EDTR à ce moment-là. Quoi qu'il en soit, dès que les concepts géographiques du Recensement de 2001 furent disponibles, les modifications se sont révélées nécessaires. Il y avait non seulement des variations cumulatives des chiffres de population sur dix ans, mais aussi des changements des limites qui étaient plus nombreux au cours de la deuxième moitié (de 1996 à 2001) qu'au cours de la première moitié de la décennie.

Afin d'illustrer ce deuxième effet, le tableau 1 ci-après montre l'incidence des modifications des limites administratives sur la population de 15 grandes RMR, de 1991 à 1996, de 1996 à 2001, et de 1991 à 2001.

De 1991 à 1996, les chiffres de population ont peu varié en raison de l'ajustement des limites. Sur les 15 RMR dont les données ont été publiées dans les tableaux normalisés de l'EDTR, seulement trois ont vu leur population varier de 1 % ou plus à

2. Étant la seule entrée de collecte aux fins du codage géographique, le code postal comporte un inconvénient : il ne représente pas toujours le lieu de résidence du ménage. Tel est le cas en particulier dans les régions rurales, où le service rural et le ramassage au bureau de poste sont fréquemment utilisés dans la livraison du courrier. Mais, dans la plupart des cas, le code postal correspond au lieu de résidence du répondant.

la suite des modifications de leurs limites. De 1996 à 2001, seulement quatre RMR ont fait l'objet de modifications aussi considérables. Cependant, de 1991 à 2001, six des 15 RMR ont été touchées, ce qui a donné à penser qu'une modification était nécessaire.

Tableau 1 : Variation des chiffres de population des RMR due à la modification de leurs limites

Nom de la RMR	Variation des chiffres de population due à la modification des limites d'un recensement à l'autre (en %)		
	1991 à 1996	1996 à 2001	1991 à 2001
Montréal	2,7 %	0,0 %	2,8 %
Ottawa-Gatineau	2,4 %	-1,2 %	1,3 %
Winnipeg	1,3 %	0,0 %	1,3 %
Halifax	0,0 %	3,3 %	3,5 %
Québec	0,0 %	0,0 %	0,0 %
Toronto	0,2 %	0,0 %	0,2 %
Hamilton	0,0 %	0,0 %	0,0 %
St. Catherines-Niagara	0,0 %	0,0 %	0,0 %
Kitchener	0,0 %	0,0 %	0,0 %
London	0,0 %	4,7 %	4,8 %
Windsor	0,0 %	3,1 %	2,9 %
Calgary	0,0 %	0,0 %	0,0 %
Edmonton	0,1 %	0,0 %	0,2 %
Vancouver	0,0 %	0,0 %	0,0 %
Victoria	0,0 %	0,0 %	0,0 %

Notes :

Il est possible que la somme des pourcentages du tableau ne corresponde pas au total en raison de l'arrondissement. Les valeurs 0,0 % du tableau sous-entendent qu'il n'y a pas eu de modification intercensitaire des limites, ou que les chiffres de population n'ont pas été considérables malgré la modification des limites.

Source : GéoRéf 91 (n° 92-345D au catalogue), GéoSuite 2001 (n° 92F0085XCB au catalogue), et la Division de la géographie.

Chevauchement des concepts géographiques du recensement dans les données de l'EDTR

Au cours du passage aux concepts géographiques du Recensement de 2001, l'équipe de l'EDTR a décidé de procéder à une révision chronologique des données infra-provinciales de 1999, 2000 et 2001, conformément aux limites géographiques des concepts géographiques du Recensement de 2001, tout en conservant également les valeurs antérieures dans un ensemble de variables distinct. Il en résulte un chevauchement de trois ans qui peut être examiné à l'aide des limites géographiques du Recensement de 1991 ou de celui de 2001. Ce choix permet l'analyse des données à l'aide d'une conception géographique uniforme pour les six années complètes de chaque panel de répondants. Le tableau 2 montre comment la base de données est structurée.

Tableau 2 : Structure des variables géographiques

Années de référence :	1992 à 1998	1999 à 2001	Depuis 2002
Variables sans « X » p. ex., CMACA25	Désigne les limites de 1991	Désigne les limites de 2001	
Variables avec « X » p. ex., XCMACA25	Désigne les limites de 1991		Codées sous la mention « Non applicable »

Toutes les valeurs géographiques des concepts géographiques du Recensement de 1991, allant de 1992 à 2001, sont stockées dans les variables qui portent les mêmes noms, sauf qu'elles débutent par un « X ». Par exemple, les données de CMACA25 ont été copiées dans XCMACA25. Les variables géographiques courantes comportent les valeurs des concepts géographiques du Recensement de 2001 pour la période qui commence en 1999.

La plupart des variables géographiques désignent le lieu du ménage à la fin de l'année de référence. Mais il y a quatre variables sur la mobilité géographique qui sont directement touchées par la modification apportée aux concepts géographiques du Recensement de 2001 : SAMCMA26 – pour les personnes habitant une RMR/AR à la fin de l'année de référence, il s'agit d'un indicateur qui précise si la personne habite la même RMR/AR qu'à la fin de l'année précédente; SAMCSD26 – indicateur qui précise si la personne habite la même SDR qu'à la fin de l'année précédente; DISTM26 – la distance parcourue exprimée en kilomètres; et MOVST29 – la situation de déménagement de la personne pour l'année de référence.

Ces quatre variables sont dérivées des autres variables géographiques au moyen de la comparaison des valeurs antérieures et actuelles des années de référence. Puisqu'il faut des données sur deux ans pour dériver la valeur d'une année donnée, on utilise maintenant les concepts géographiques du Recensement de 2001 depuis 2000 (ce qui constitue la comparaison entre 1999 et 2000). Ainsi, pour les années de référence antérieures à 2000, on utilise les concepts géographiques du Recensement de 1991.

Concepts géographiques normalisés nouveaux et abandonnés

Les concepts géographiques du Recensement de 2001 comprennent trois nouveaux concepts géographiques normalisés qui ne figuraient pas dans les recensements de 1991 et de 1996 (ni pour les années antérieures).

1. **L'aire de diffusion (AD)³** est une unité géographique restreinte et relativement stable composée d'un îlot ou plus. Depuis 2001, l'aire de diffusion est la zone géographique normalisée la plus petite pour laquelle on puisse diffuser des données de recensement (c.-à-d. mises en tableau à des fins d'analyse), d'où le nom qu'on lui a attribué. Par le passé, la plus petite zone

3. Pour en savoir plus sur ces concepts, reportez-vous au document : Division des opérations du recensement, *Dictionnaire du recensement de 2001, Référence*, n° 92-378-XPf au catalogue de Statistique Canada, 2002.

de diffusion des données était le **secteur de dénombrement (SD)**⁴. Le SD est une zone géographique visitée par un recenseur et qui sert toujours à des fins de collecte.

2. La **zone d'influence des régions métropolitaines de recensement et des agglomérations de recensement (ZIM)**⁵ est un nouveau concept qui a été introduit au moment de l'élaboration des concepts géographiques du Recensement de 2001. Ce concept permet de classer les subdivisions de recensement (SDR) qui ne font partie ni d'une RMR ni d'une AR, selon le degré d'influence que les RMR ou les AR exercent sur ces subdivisions. Le degré d'influence sur la SDR est estimé en fonction de l'étendue de la navette entre cette SDR et les autres zones – plus précisément, la proportion de la population active occupée totale qui habite la SDR et qui travaille dans un noyau urbain de RMR ou d'AR. Les SDR où le navettage est égal ou supérieur à 30 % sont classées dans les ZIM « fortes ». Celles où le navettage est égal ou supérieur à 5 % mais inférieur à 30 % sont appelées ZIM « modérées ». Puis les SDR où le navettage est supérieur à 0 % mais inférieur à 5 % sont comprises dans la catégorie des ZIM « faibles ». Enfin, les SDR dont la population active résidente qui fait la navette pour se rendre au travail dans des noyaux urbains de RMR ou d'AR est inférieure à 40 personnes sont classées dans les ZIM « sans influence ».
3. La **classification des secteurs statistiques (CSS)**⁶ permet de grouper les subdivisions de recensement (SDR) selon qu'elles font partie d'une RMR, d'une AR, d'une ZIM ou des territoires. Cette classification davantage détaillée remplace celle de **type RMR/AR**. Les RMR et les AR ne sont plus classées comme étant « unifiées » ou « primaires ». De plus, les AR (mais pas les RMR) sont classées comme étant « divisées en SR » ou « non divisées en SR ». Enfin, toutes les autres zones qui ne figuraient dans aucune classification auparavant figurent maintenant dans les ZIM.

Le tableau 3 comporte les variables géographiques figurant dans la base de données de l'EDTR. Les deux premières colonnes donnent les variables géographiques de l'EDTR qui font référence aux concepts géographiques du Recensement de 1991. La troisième colonne indique si ces mêmes variables des concepts géographiques du Recensement de 1991 figurent aussi dans les concepts géographiques du Recensement de 2001.

4. Pour en savoir plus sur ces concepts, reportez-vous au document : Division des opérations du recensement, *Dictionnaire du recensement de 1991, Référence*, n° 92-301F au catalogue de Statistique Canada, 1992.

5. Pour en savoir plus sur ces concepts, reportez-vous au document : Division des opérations du recensement, *Dictionnaire du recensement de 2001, Référence*, n° 92-378-XPF au catalogue de Statistique Canada, 2002.

6. Pour en savoir plus sur ces concepts, reportez-vous au document : Division des opérations du recensement, *Dictionnaire du recensement de 2001, Référence*, n° 92-378-XPF au catalogue de Statistique Canada, 2002.

Tableau 3 : Variables géographiques de la base de données de l'EDTR

Variable de 1991	Description	Dans les concepts géographiques du Recensement de 2001
PVRES25	Province de résidence	Oui
PVREG25	Groupe de province de résidence	Oui
FEDRES25	Circonscription électorale fédérale ⁷	Oui
EARES25	Secteur de dénombrement	Non, remplacée par l'aire de diffusion : DARES25
CDRES25	Division de recensement	Oui
CSDRES25	Subdivision de recensement	Oui
REGRES25	Région	Oui
ERRES25	Région économique	Oui
CMACA25	RMR/AR	Oui
CMA1G25	RMR/AR du groupe 1	Oui
CMA2G25	RMR/AR du groupe 2	Oui
CMAPO25	Population de RMR/AR	Oui
CATYP25	Type de RMR/AR	Non, remplacée par le type de classification des régions normalisées : SACTYP25 (comprend les classifications des ZIM)
LNGRES25	Longitude de résidence	Oui
LATRES25	Latitude de résidence	Oui
URBRUR25	Ménage urbain/rural	Oui
URBCD25	Code de région urbaine	Oui
URBPO25	Population du secteur de résidence	Oui
URBSZ25	Taille urbaine	Oui
URBSZG25	Taille du secteur de résidence	Oui
USZGA25	Taille du secteur de résidence ajustée	Oui
EIR25	Région d'assurance-emploi ⁸	Oui
POSTCD25	Code postal	Oui
POSTCI25	Indicateur d'imputation du code postal	Oui

Fusion des deux concepts géographiques en une série de variables géographiques

Les concepts géographiques qui sont communs aux recensements de 1991 et de 2001 continuent d'exister sous les mêmes noms de variables dans l'EDTR, tout comme avant la diffusion des données de 2002. Au lieu de créer de nouvelles variables,

7. Les CÉF sont représentées par un député de la Chambre des communes.

8. Pour les concepts géographiques du Recensement de 1991, on utilise la délimitation de juillet 1996 des régions d'assurance-emploi. Pour les concepts géographiques du Recensement de 2001, on utilise la délimitation de juillet 2000 des régions d'assurance-emploi.

l'équipe de l'EDTR a augmenté les catégories et descriptions au fur et à mesure, afin d'intégrer les régions et les nouveaux noms.

Les catégories géographiques qui ne s'appliquent qu'à certaines années (parce qu'elles ne s'appliquent qu'à l'un des deux concepts géographiques de recensement) apparaîtront encore dans la série de codes, mais comporteront une remarque précisant les années pour lesquelles elles ne s'appliquent pas. Pour ces années-là, il n'y aura pas de ménages portant ces codes.

Les trois exemples ci-après illustrent comment la variable CMACA25 a été modifiée pour représenter les deux concepts géographiques. Il convient de noter qu'une municipalité doit satisfaire à certains critères de population pour être admise comme région métropolitaine de recensement (RMR) ou agglomération de recensement (AR); les AR peuvent être créées ou retirées d'un recensement à l'autre, et si les RMR ne sont jamais retirées, de nouvelles peuvent être créées. On donne aux ménages des régions autres que les RMR/AR la valeur « Non Applicable » à la variable CMACA25.

Cas A : Les codes qui ne sont valides qu'avant l'année 1999 (selon les concepts géographiques du Recensement de 2001). Le code 730 de la variable CMACA25 représente Weyburn en Saskatchewan. Il s'agissait d'une AR dans les concepts géographiques du Recensement de 1991, titre qui a toutefois été retiré en 1996 parce que la population a chuté sous la barre des 10 000 habitants selon les résultats du recensement (elle a conservé son statut de région autre que RMR/AR en 2001). Les ménages de Weyburn ne font plus partie d'une RMR ou AR; on a donc attribué pour ces ménages et pour la variable CMACA25 le code « Non Applicable » depuis 1999. Un exemple de RMR/AR disparue à cause d'une modification des limites géographiques est Saint-Jérôme, au Québec, représentée par le code 475. Il s'agissait d'une AR autonome dans les concepts géographiques du Recensement de 1991, mais elle a été fusionnée à Montréal, laquelle comptait déjà sa propre RMR avant le Recensement de 2001. Contrairement à ceux de Weyburn, les ménages de Saint-Jérôme font encore partie d'une RMR/AR, et on leur attribue le code de Montréal depuis 1999.

Cas B : Les codes qui ne s'appliquent pas avant 1999 (selon les concepts géographiques du Recensement de 2001). Un exemple d'AR qui a reçu ce titre dans les concepts géographiques du Recensement de 2001 mais pas dans celle de 1991 est Squamish, en Colombie-Britannique. À cette fin, un nouveau code a été créé pour la variable CMACA25, il s'agit du code 934.

Cas C : Les codes qui s'appliquent pour toutes les années de référence. La plupart des codes ont conservé le même nom et s'appliquent encore parce que la région demeure une RMR ou une AR autonome. De plus, les limites peuvent avoir changé s'il y a eu modification des secteurs de compétence municipale; cela n'est pas indiqué dans la variable CMACA25, mais est pris en considération dans la cartographie particulière au recensement des SDR aux RMR, des AR, et des régions autres que les RMR/AR. Le code 225 de la variable CMACA25, qui portait le nom de Sydney, en

Nouvelle-Écosse, dans les concepts géographiques du Recensement de 1991, a été renommé Cap Breton en raison de la fusion et de la restructuration municipale. Voilà un exemple de modification de nom conjuguée à une modification des limites. La description liée à ce code indique que celui-ci a représenté Sydney jusqu'en 1998, après quoi il a représenté Cap Breton.

Comment procède-t-on au traitement géographique dans l'EDTR?

La présente section décrit les fichiers d'entrée et les étapes de traitement suivies par l'EDTR pour 2002 et rétroactivement pour 1999, 2000 et 2001 dans les données révisées⁹.

Fichiers d'entrée

Il y a quatre principaux fichiers d'entrée qui sont nécessaires au traitement des données géographiques de chaque ménage. Ces fichiers sont : les Fichiers de conversion des codes postaux (FCCP); le Fichier des attributs géographiques (FAG); le fichier de la région d'assurance-emploi (RAE); et le fichier de facturation du service téléphonique.

Le **Fichier de conversion des codes postaux (FCCP)** produit par la Division de la géographie comporte le lien entre les codes postaux à six caractères et les régions géographiques normalisées du recensement de Statistique Canada (comme les AD, les municipalités et les RMR ou AR). De plus, ce fichier fournit les coordonnées de latitude et de longitude d'un point représentant le lieu approximatif du centre de la région définie selon le code postal. Outre une mise à jour quinquennale visant à refléter la nouvelle géographie du recensement, le FCCP est mis à jour tous les six mois pour tenir compte des modifications continuellement apportées aux codes postaux par la Société canadienne des postes (SCP)¹⁰.

Le **Fichier des attributs géographiques (FAG)** comporte les chiffres de la population et le nombre de logements du dernier recensement, selon l'aire de diffusion (AD). Il fournit aussi la correspondance ou le « lien » entre chacune des AD et tous les niveaux géographiques supérieurs, sauf la région d'assurance-emploi. Le FAG comprend les données qui déterminent les régions urbaines et rurales (indicateur de région urbaine ou rurale, code de région urbaine et population de région urbaine), ainsi que la longitude et la latitude d'un point représentatif de chaque AD.

9. Pour des renseignements sur la manière dont l'information géographique a été traitée dans l'EDTR à l'aide la conception du Recensement de 1991, voir *Les données de l'EDTR relatives à la géographie et leur incidence sur les mesures de faible revenu* par R. Cunningham, P. Lafrance, J. Rowland et J. Murray, Série de documents de travail sur la dynamique du travail et du revenu, n° 97-09 au catalogue de Statistique Canada, 1991.

10. Le FCCP a d'abord été couplé aux régions géographiques du Recensement de 1981, année depuis laquelle il a fait l'objet de quatre « conversions » afin de refléter les recensements de 1986, de 1991, de 1996 et de 2001. Pour en savoir plus, voir *Le fichier de conversion des codes postaux (PCCF). guide de référence*, n° 95F0153G1F au catalogue de Statistique Canada.

Le fichier de la **Région d'assurance-emploi (RAE)** fournit le lien entre chaque AD et les régions d'assurance-emploi (RAE), lesquelles sont définies par Ressources humaines et Développement des compétences Canada (RHDCC) et Développement social Canada (DS) (anciennement Développement des ressources humaines Canada). L'admissibilité aux prestations d'assurance-emploi est fondée en partie sur le taux de chômage de la RAE du demandeur.

Les trois fichiers susmentionnés (FCCP, FAG et RAE) peuvent être fusionnés en un seul fichier, à l'aide de l'aire de diffusion comme champ commun à ces fichiers. Que les données de l'EDTR soient couplées à ces fichiers individuellement ou à la suite de la fusion des trois fichiers importe peu. Pour simplifier la description des étapes de traitement ci-après, nous pouvons les désigner conjointement sous le nom de fichier FCCP-FAG-RAE.

L'équipe de l'EDTR produit chaque année un seul **fichier de facturation du service téléphonique**, à partir de nombreux fichiers de facturation téléphonique qu'ils reçoivent trimestriellement d'environ 22 entreprises fournissant des services téléphoniques dans l'ensemble du Canada. À noter que ces fichiers de facturation téléphonique ne comportent ni les noms des abonnés ni les renseignements relatifs à la facturation. Le numéro de téléphone, le code postal et les champs provinciaux suffisent au traitement de l'EDTR¹¹.

Une fois ces fichiers obtenus, on peut procéder au traitement géographique de l'EDTR pour chaque ménage.

Étapes d'affectation des valeurs géographiques

1^{re} étape

Après avoir mis de côté tous les enregistrements sans code postal pour l'année en cours (ils seront traités dans la troisième étape), on procède à un appariement des ménages de l'année en cours avec ceux de l'année précédente, selon la variable d'identification du ménage, la province et le code postal. Pour tous les ménages de l'année en cours qui font l'objet d'un appariement, on reproduit les codes géographiques de recensement de l'année précédente.

Pour ce qui est des enregistrements non appariés, avant de passer à l'étape suivante, on vérifie la cohérence du code postal et du numéro de téléphone avec le code de province. Par exemple, les ménages qui habitent l'Ontario doivent avoir un code postal qui commence par les lettres K, L, M, N ou P, et l'indicatif régional du numéro

11. Malgré leur utilité, les fichiers de facturation téléphonique comportent quelques inconvénients qui peuvent empêcher d'établir le lien avec les répondants de l'enquête, ou qui peuvent indiquer une adresse autre que celle de la résidence permanente de ces derniers. Il est possible que la facturation soit postée à une résidence ou un bureau de poste autre, quoique cela importe peu en réalité sauf si le code postal de l'adresse diffère de celui de la résidence. Selon l'entreprise fournissant le service téléphonique, il se peut que les numéros confidentiels ne soient pas inclus, ce qui empêchera l'appariement.

de téléphone sera nécessairement 289, 416, 519, 613, 647, 705, 807 ou 905. Si le code postal ou le numéro de téléphone n'est pas conforme à cette vérification, ils sont alors désignés comme « manquants » et seule la variable « province » est conservée. Ou bien on assignera des valeurs aux variables géographiques à l'étape 3, ou bien il y aura imputation.

2^e étape

Appariement des enregistrements non appariés à la 1^{re} étape avec le fichier FCCP-FAG-RAE par province et code postal. Tous les enregistrements comportant un code postal seront appariés au fichier, à moins que certains codes non valides n'aient pas été saisis à la 1^{re} étape. Pour simplifier le processus, nous pouvons supposer que tous les enregistrements non appariés à cette étape-ci ont la valeur « manquant » pour le code postal.

3^e étape

Appariement des enregistrements non appariés lors de la 2^e étape et des enregistrements sans code postal (mis de côté dans la 1^{re} étape) au fichier de facturation du service téléphonique, par numéro de téléphone à dix chiffres. Afin de réussir l'appariement des enregistrements, le code postal est ajouté puis l'enregistrement est apparié de nouveau au fichier FCCP-FAG-RAE pour obtenir toutes les valeurs géographiques. Les enregistrements pour lesquels cette procédure d'appariement ne fonctionne pas passent par la 4^e étape.

4^e étape

Les enregistrements qui ne comportent pas encore de code postal après la 3^e étape sont comparés au fichier de l'année précédente à la même étape (c.-à-d. tout juste après la 3^e étape) selon la variable d'identification du ménage, l'adresse à l'exclusion du code postal et du numéro de téléphone. Si les enregistrements sont les mêmes, les valeurs géographiques sont alors reportées à l'année en cours afin d'éviter une modification du lieu de résidence par réimputation. Les autres enregistrements auront un code postal assigné par imputation.

5^e étape

Imputation des enregistrements non appariés à la 4^e étape. Veuillez consulter la prochaine section « méthodes d'imputation ». Une fois le code postal, la DR et l'AD imputés, nous apparions ces enregistrements au fichier FCCP-FAG-RAE pour retrouver les autres valeurs géographiques.

Le traitement géographique est alors complet, sauf pour certaines variables qui ne sont pas directement fournies par le fichier FCCP-FAG-RAE mais qui peuvent être dérivées, telles que la distance parcourue.

Le tableau 4 donne les chiffres des enregistrements relatifs aux étapes de traitement qui figurent ci-dessus.

Tableau 4 : Codes géographiques assignés aux enregistrements, selon la méthode (selon les concepts géographiques du Recensement de 2001)

Année de référence	1999	2000	2001	2002
Nombre total de ménages dans l'échantillon	35 761 (100 %)	37 125 (100 %)	36 444 (100 %)	36 766 (100 %)
Enregistrements à l'extérieur du Canada : La mention « Non Applicable » apparaît dans tous les champs	305 (1 %)	389 (1 %)	481 (1 %)	307 (1 %)
1 ^{re} étape : Codes reproduits à partir de l'année précédente, puisque le code postal et la variable d'identification du ménage sont identiques	*	25 491 (69 %)	27 572 (76 %)	14 646 (40 %)
2 ^e étape : Codes obtenus par l'appariement du code postal au fichier FCCP-FAG-RAE, puisque le code postal est déjà fourni et valide	34 064 (95 %)	3 585 (10 %)	6 084 (17 %)	19 875 (54 %)
3 ^e étape : Codes obtenus au moyen du numéro de téléphone pour obtenir le code postal, puis appariés au fichier FCCP-FAG-RAE	800 (2 %)	4 916 (13 %)	728 (2 %)	860 (2 %)
4 ^e étape : Codes reproduits à partir des valeurs de l'année précédente puisque le code postal est manquant, mais les adresses et la variable d'identification du ménage sont identiques	*	81 (0 %)	180 (0 %)	247 (1 %)
5 ^e étape : Le code postal, la DR et l'AD sont imputés, après quoi l'enregistrement est apparié au fichier FCCP-FAG-RAE	592 (2 %)	2 663 (7 %)	1 399 (4 %)	831 (2 %)

* « Ne s'applique pas » puisque 1999 est la toute première année faisant l'objet du traitement à l'aide des concepts géographiques du Recensement de 2001.

Nota : En 2002, dans la première étape, le pourcentage d'enregistrements ayant des codes géographiques reproduits de l'année précédente est inférieur à celui de 2000 et 2001 en raison de l'insertion d'un nouveau panel en 2002.

Méthodes d'imputation

Les méthodes d'imputation par code postal figurent ci-dessous selon l'ordre dans lequel elles servent au traitement de l'EDTR. Ces méthodes ont été utilisées chaque année depuis le traitement des données de 2002, mais elles s'appliquent aussi aux années 1999, 2000 et 2001 parce que les données géographiques imputées de ces années ont été réimputées dans une révision chronologique. Avant le traitement des données de 2002, on se servait de la méthode très simple de tirage au hasard d'un des codes postaux de la province.

Toutes les méthodes d'imputation décrites ici se trouvent au sein d'une province donnée. Une des raisons principales pour lesquelles l'imputation se limite au sein de la province est le respect de la procédure de pondération des révisions chronologiques de 1999 à 2001. Une autre raison est la nécessité d'imputer la province au préalable pour certains enregistrements, puisque le personnel de l'EDTR n'a pu les retracer. La

méthode utilisée pour l'imputation de la province ne consiste qu'à supposer que le ménage habite encore la même province que la dernière province connue.¹²

1. Imputation longitudinale – rétrospective (POSTCI25=40)

- Ce type d'imputation ne peut servir qu'aux révisions chronologiques, puisqu'il consiste à prendre les renseignements connus d'une année puis les appliquer à une année antérieure ou plus.
- Cette méthode a été utilisée pour les années ayant fait l'objet d'une révision chronologique, de 1999 à 2001.
- Si l'enregistrement d'un ménage se trouve dans l'enquête pendant au moins deux années consécutives et que l'adresse à la deuxième année est connue, mais qu'elle est inconnue à la première année, on suppose alors que le personnel de l'EDTR a perdu contact avec le répondant lors de la première année des données manquantes. Cette supposition est logique puisque souvent l'impossibilité, pour le personnel de l'EDTR, de retracer un répondant est attribuable au déménagement de celui-ci. On suppose alors que le répondant n'a pas déménagé entre les deux années en question, de sorte que l'adresse de l'année la plus récente soit imputée à l'année antérieure¹³.

2. Imputation longitudinale – prospective (POSTCI25=43)

- Cette imputation ne sert qu'aux années de référence depuis 2000 – après avoir traité de nouveau les données de 1999.
- S'il faut imputer le code postal du ménage pour deux années consécutives, ce ménage reçoit alors le même code postal pour ces deux années. Cette pratique a lieu parce qu'on n'a pas de renseignements indiquant si le ménage a déménagé ou pas. Le personnel de l'EDTR n'a pas voulu imputer de nouveau le code postal de ce ménage et lui attribuer de nouveau un lieu. Cela suppose nécessairement qu'il n'y a pas eu de déménagement.

3. Imputation à l'aide du numéro de téléphone à six chiffres (indicatif régional + banque d'échange) (POSTCI25=41)

- Pour les enregistrements qui comportent un numéro de téléphone pour l'année de référence actuelle, tirez les six premiers chiffres (indicatif régional + banque d'échange). Par exemple, les six premiers chiffres du numéro de téléphone d'un résident d'Ottawa pourraient être (613744), où (613) est l'indicatif régional d'Ottawa et (744), cela représente la banque d'échange. Nous apparions ces deux données aux numéros de téléphone à

12. L'imputation du lieu du ménage au sein de la province de la dernière adresse connue est une limite artificielle, puisqu'il est possible en réalité que les répondants aient déménagé dans une autre province. L'idéal serait que la méthode d'imputation permette les déplacements interprovinciaux. Mais puisque cela exige énormément de recherche dans différentes méthodes avant sa mise en œuvre, elle a été exclue des récents changements apportés au traitement géographique, lesquels comprenaient déjà l'application des nouveaux concepts du recensement, des méthodes d'imputation améliorée, et des révisions chronologiques. Cependant, on a l'intention de l'envisager dans un avenir prochain.

13. Les codes postaux n'ont pas été touchés par la modification des concepts géographiques.

six chiffres du fichier de facturation du service téléphonique, lequel a d'abord été fusionné avec le FCCP. S'il y a de nombreux codes postaux pour cet indicatif régional et cette banque d'échange, nous en tirons un au hasard.

- Si dans l'année de référence actuelle le numéro de téléphone ne peut être déterminé ou s'il n'y a pas de numéro de téléphone pour le répondant, on utilise alors la méthode 4.

4. Imputation de la distance à l'aide du code postal de l'année précédente (POSTCI25=42)

- Cette méthode est utilisée pour les enregistrements qui se trouvaient dans l'enquête l'année précédente et qui ne devaient pas être imputés cette année-là; on connaît donc avec certitude leur lieu d'il y a un an. La première étape de cette méthode est d'imputer la distance parcourue. On suppose que les ménages ont déménagé parce que le personnel de l'EDTR n'a pu les retracer pour l'enquête. À la suite de cette première étape, on connaît toutes les coordonnées longitudinales et latitudinales de leurs lieux de déménagement possibles, coordonnées qui correspondront à la distance de déménagement parcourue. La deuxième étape consiste à choisir une de ces paires de coordonnées et à s'en servir pour obtenir le code postal. On trouvera ci-dessous quelques autres détails sur la façon de faire l'imputation selon la distance.
- On calcule la « distance parcourue » pour chaque répondant de l'EDTR figurant dans l'enquête des années actuelle et antérieure, dont le code postal est valide pour ces deux années (c.-à-d. sans imputation) et dont celui-ci a changé. Toutes ces valeurs sont triées par province et type de déménagement (c.-à-d. urbain-urbain, urbain-rural, rural-urbain ou rural-rural). Ensuite, pour chaque ménage nécessitant une imputation, on tire au hasard une des valeurs de la « distance parcourue » à partir de celles au sein de la même province et du même type de déménagement que le ménage. Il s'agira de l'enregistrement donneur qui sert à établir la distance approximative parcourue par le ménage nécessitant l'imputation.
- Le FCCP sert à obtenir la liste des coordonnées longitudinales et latitudinales de toutes les régions définies selon le code postal dans la province; il ne donne qu'une paire de coordonnées pour chaque région définie selon le code postal. Ensuite, pour chaque ménage nécessitant l'imputation, on doit trouver toutes les destinations possibles où les ménages ont pu déménager au sein de la même province, compte tenu de la distance parcourue et du lieu antérieur. Pour obtenir un certain nombre de destinations, on étend la distance parcourue du donneur dans un rayon de plus ou moins cinq kilomètres (par exemple) de l'enregistrement du ménage à imputer. Dans la liste des destinations trouvées, une destination est choisie au hasard et tient lieu de destination où le ménage a déménagé. On peut obtenir ensuite le code postal à partir des coordonnées.

5. Imputation au hasard du code postal/de l'AD/de la DR (POSTCI25=44)

- On n'applique cette méthode de dernier recours que si le code postal n'a pu être obtenu à l'aide des méthodes susmentionnées. Le code postal au sein d'une province est choisi au hasard.
- Les enregistrements qui nécessitent l'imputation au hasard ne comportent habituellement ni code postal ni numéro de téléphone. Dans le cas de nombreux enregistrements, il n'y a que la variable d'identification du ménage et la province de résidence, les autres champs d'information étant vides.

La variable POSTCI25 sert à consigner la façon dont le code postal final a été obtenu pour chacun des enregistrements des ménages. Les diverses catégories figurent dans le tableau 5 ci-dessous. Parmi les catégories possibles, il y a les différents types de dérivation ou d'imputation qui ont été appliqués, selon l'année. Les codes 30, 31 et 32 ne s'appliquent maintenant qu'aux années de référence 1992 à 1998, lesquelles sont encore basées sur la géographie du Recensement de 1991 dans l'EDTR. Les codes 40 à 44 font référence à la géographie du Recensement de 2001 et aux méthodes d'imputation mises à jour de l'EDTR. Le code 40 ne s'applique que de 1999 à 2001, tandis que le code 43 ne s'applique qu'à partir de 2000. En outre, les codes 41, 42 et 44 ne s'appliquent qu'à partir de 1999.

Tableau 5 : Catégories de variable POSTCI25 : Comment le code postal a été assigné

Code	Description
10	Code postal recueilli
20	Code postal dérivé du numéro de téléphone
30	Circonscription électorale fédérale/Secteur de dénombrement assignés par la Division de la géographie à l'aide de l'adresse (AR1992-AR1998)
31	Circonscription électorale fédérale/Secteur de dénombrement imputés par la Division de la géographie à l'aide de la province (AR1992-AR1998)
32	Circonscription électorale fédérale/Secteur de dénombrement assignés à partir de la strate – du type – de la grappe de la base de sondage de l'EPA (AR1992-AR1993)
40	Imputation longitudinale – rétrospective (AR1999-AR2001)
41	Imputation à l'aide du numéro de téléphone à six chiffres de l'année de référence (indicatif régional + banque d'échange) (c.-à-d. 613-951) (depuis l'AR1999)
42	Imputation de la distance à l'aide du code postal de l'année précédente (depuis l'AR1999)
43	Imputation longitudinale – prospective (depuis l'AR2000)
44	Imputation au hasard du code postal/de l'AD/de la DR (depuis l'AR1999)

Nota : AR désigne l'année de référence. Les années de référence figurant entre parenthèses signifient que le code ne sert qu'à ces années-là.

Le tableau 6 comporte le nombre d'enregistrements dont les codes postaux ont été imputés par des méthodes différentes.

Tableau 6 : Chiffres selon les méthodes d'imputation (POSTCI25)

Code	Méthodes d'imputation	1999	2000	2001	2002
40	Imputation longitudinale – rétrospective (AR1999-AR2001)	41 (6,9 %)	1 031 (38,7 %)	40 (2,9 %)	...
41	Imputation à l'aide du numéro de téléphone à six chiffres de l'année de référence (indicatif régional + banque d'échange, p. ex. 613-951) (depuis l'AR1999)	187 (31,6 %)	793 (29,8 %)	146 (10,4 %)	179 (21,5 %)
42	Imputation de la distance à l'aide du code postal de l'année précédente (depuis l'AR1999)	303 (51,2 %)	751 (28,2 %)	982 (70,2 %)	483 (58,1 %)
43	Imputation longitudinale – prospective (depuis l'AR2000)	...	0 (0,0 %)	151 (10,8 %)	20 (2,4 %)
44	Imputation au hasard du code postal/de l'AD/de la DR (depuis l'AR1999)	61 (10,3 %)	88 (3,3 %)	80 (5,7 %)	149 (17,9 %)
...	Total	592 (100,0 %)	2 663 (100,0 %)	1 399 (100,0 %)	831 (100,0 %)

Nota : AR désigne l'année de référence. Les années de référence figurant entre parenthèses signifient que le code ne sert qu'à ces années-là.

En appliquant ces méthodes améliorées pour remplacer les valeurs manquantes des renseignements géographiques, le personnel de l'EDTR a remarqué une amélioration supplémentaire qui pourrait être apportée.

La méthode d'imputation de la distance décrite ci-dessus pour les années 1999 à 2002 est basée sur une formule approximative :

$$(1) DM = 75 * \text{SQRT}[(LATRES_{RY} - LATRES_{RY-1})^2 + (LNGRES_{RY} - LNGRES_{RY-1})^2],$$

où

- DM = distance parcourue
- SQRT = racine carrée
- LATRES = latitude de résidence
- LNGRES = longitude de résidence
- RY = année de référence actuelle
- RY-1 = année de référence antérieure

La formule (1) a tendance à sous-estimer la distance parcourue d'environ 33 % pour les latitudes inférieures, et à la surestimer d'environ 20 % pour les latitudes supérieures.

À partir de 2003, elle a été remplacée par une formule plus précise :

$$(2) DM = 111.32 * \arccos[\sin(LATRES_{RY}) * \sin(LATRES_{RY-1}) + \cos(LATRES_{RY}) * \cos(LATRES_{RY-1}) * \cos(LNGRES_{RY} - LNGRES_{RY-1})]$$

La formule (2) calcule la distance le long d'un « grand arc circulaire », c.-à-d. la distance de surface la plus courte entre deux points d'une sphère (c.-à-d. en supposant que la Terre est une sphère parfaite).

Passage aux concepts géographiques du Recensement de 2001 : l'effet sur les données

Les données antérieures à 1999 sont basées sur les concepts géographiques du Recensement de 1991. Comme les données des années de référence depuis 1999 sont basées sur les concepts géographiques du Recensement de 2001, il y a une rupture structurelle dans les séries de données de 1998 à 1999. Puisque bon nombre de limites géographiques ont été modifiées, tout comme les noms et les codes, il se peut que les codes géographiques de certains ménages diffèrent entre 1998 et 1999 même s'il n'y a pas eu de déménagement.

De plus, en raison des modifications apportées aux méthodes d'imputation, il se peut que certains ménages imputés aux années de référence 1999, 2000 ou 2001 se trouvent à un autre lieu au moment du passage aux différents concepts géographiques des recensements de 1991 à 2001.

La mise à jour des renseignements géographiques peut influencer sur les taux de faible revenu en particulier. Par exemple, la taille de la famille et la taille de la collectivité servent à trouver le seuil adéquat, après quoi le revenu familial est comparé à ce seuil. Alors, si l'on calcule le taux de faible revenu des familles, celles-ci sont recensées comme étant à faible revenu si leur revenu est inférieur à ce seuil. Si l'on calcule le faible revenu d'un individu, tous les individus membres de la famille sont recensés comme étant à faible revenu si le revenu familial est inférieur au seuil.

Les seuils de faible revenu (SFR) sont établis à partir des données de l'Enquête sur les dépenses des familles, rebaptisée *Enquête sur les dépenses des ménages*. Les SFR désignent le niveau de revenu auquel une famille risque de se trouver en difficulté parce qu'elle doit consacrer aux nécessités une part plus importante de son revenu que la famille moyenne de taille semblable. De façon précise, on définit le seuil comme étant le revenu en deçà duquel une famille est susceptible de dépenser 20 points de pourcentage de plus de son revenu que la famille moyenne. Il existe des seuils distincts pour sept tailles de famille – allant des personnes hors famille aux familles d'au moins sept personnes – et pour cinq tailles de collectivité – allant des zones rurales aux zones urbaines de plus de 500 000 habitants.

Le tableau 7 ci-dessous donne les taux de faible revenu (la proportion de personnes à faible revenu) des années de référence 1999, 2000 et 2001 pour les conceptions de recensement de 1991 et de 2001. Ce tableau indique que le passage aux concepts géographiques du Recensement de 1991 à celle du Recensement de 2001 visant à traiter les variables géographiques n'a pas influé considérablement sur les taux de faible revenu.

Tableau 7 : Taux de faible revenu dans 15 grandes RMR au Canada

Conception du recensement	1999			
	Avant impôt		Après impôt	
	Recensement de 1991	Recensement de 2001	Recensement de 1991	Recensement de 2001
Montréal	23,9 %	24,1 %	19,3 %	19,4 %
Ottawa-Gatineau	17,7 %	17,9 %	14,6 %	14,7 %
Winnipeg	19,2 %	19,3 %	15,7 %	15,9 %
Halifax	17,1 %	17,1 %	13,3 %	13,5 %
Québec	18,2 %	18,6 %	14,0 %	14,4 %
Toronto	14,8 %	14,8 %	11,2 %	11,2 %
Hamilton	12,1 %	12,2 %	9,2 %	9,2 %
St. Catharines-Niagara	9,6 %	9,5 %	5,5 %	5,5 %
Kitchener	10,8 %	10,8 %	8,0 %	8,0 %
London	16,0 %	16,1 %	12,8 %	12,8 %
Windsor	12,8 %	13,4 %	8,1 %	8,3 %
Calgary	13,5 %	13,5 %	11,2 %	11,3 %
Edmonton	15,8 %	15,8 %	12,5 %	12,5 %
Vancouver	20,2 %	20,2 %	16,5 %	16,6 %
Victoria	15,4 %	15,2 %	10,6 %	10,6 %
Conception du recensement	2000			
	Avant impôt		Après impôt	
	Recensement de 1991	Recensement de 2001	Recensement de 1991	Recensement de 2001
Montréal	21,2 %	21,7 %	17,6 %	17,6 %
Ottawa-Gatineau	14,4 %	14,8 %	12,5 %	12,7 %
Winnipeg	18,3 %	18,8 %	14,8 %	15,3 %
Halifax	14,7 %	15,0 %	12,0 %	12,2 %
Québec	16,7 %	17,3 %	13,8 %	14,1 %
Toronto	13,3 %	13,4 %	10,5 %	10,7 %
Hamilton	11,5 %	11,5 %	9,0 %	9,1 %
St. Catharines-Niagara	7,5 %	7,4 %	5,5 %	5,5 %
Kitchener	10,0 %	10,0 %	8,1 %	8,1 %
London	12,0 %	12,0 %	9,8 %	9,4 %
Windsor	11,6 %	11,4 %	7,8 %	7,8 %
Calgary	12,2 %	12,2 %	8,8 %	9,0 %
Edmonton	17,0 %	17,0 %	12,2 %	12,2 %
Vancouver	17,3 %	17,3 %	14,4 %	14,5 %
Victoria	18,3 %	18,3 %	14,8 %	13,9 %

Conception du recensement	2001			
	Avant impôt		Après impôt	
	Recensement de 1991	Recensement de 2001	Recensement de 1991	Recensement de 2001
Montréal	19,0 %	19,0 %	14,8 %	14,9 %
Ottawa-Gatineau	11,4 %	11,6 %	8,1 %	8,3 %
Winnipeg	16,1 %	16,5 %	11,2 %	11,6 %
Halifax	14,2 %	14,4 %	11,7 %	11,7 %
Québec	17,2 %	17,7 %	13,1 %	13,6 %
Toronto	11,0 %	11,1 %	8,4 %	8,4 %
Hamilton	10,6 %	10,6 %	6,9 %	6,9 %
St. Catherines-Niagara	7,4 %	7,4 %	4,8 %	4,5 %
Kitchener	8,1 %	8,1 %	4,5 %	4,5 %
London	11,6 %	11,6 %	7,9 %	7,9 %
Windsor	9,4 %	9,6 %	7,9 %	7,9 %
Calgary	11,7 %	11,7 %	8,9 %	9,0 %
Edmonton	15,1 %	15,2 %	9,8 %	9,8 %
Vancouver	17,4 %	17,4 %	13,3 %	13,3 %
Victoria	14,5 %	14,5 %	10,1 %	10,1 %

Conclusion

Le personnel de l'EDTR savait qu'il serait nécessaire à un moment ou un autre d'aligner les définitions géographiques de l'EDTR à celles d'un recensement plus récent. À cette fin, les définitions géographiques de l'EDTR ont été modifiées depuis 1999, tout en chevauchant les définitions de 1999, de 2000 et de 2001. L'effet de ces modifications, comme on l'a vu dans la section précédente, est modeste et il permettra au personnel de l'EDTR une plus grande marge de manœuvre jusqu'à la prochaine modification, peut-être à la suite du Recensement de 2006 ou de 2011.

Bien que le présent document ait mis l'accent sur l'imputation, le codage géographique se fait en grande partie par la reproduction des codes de l'année précédente ou l'appariement du code postal valide du répondant aux fichiers géographiques. Les autres éléments sont traités au moyen de méthodes assez sophistiquées qui permettront à l'utilisateur de recevoir le meilleur codage géographique que le personnel de l'EDTR puisse offrir.

Ouvrages de référence

Cunningham, R., Lafrance, P., Rowland, J. et J. Murray. *Les données de l'EDTR relatives à la géographie et leur incidence sur les mesures de faible revenu*, Série de documents de travail sur la dynamique du travail et du revenu, n° 97-09 au catalogue de Statistique Canada, 1997.

Division des opérations du recensement. *Dictionnaire du recensement de 1991, Référence*, n° 92-301F au catalogue de Statistique Canada, 1992.

Division des opérations du recensement. *Dictionnaire du recensement de 1996, Édition définitive – Référence*, n° 92-351-UPF au catalogue de Statistique Canada, 1999

Division des opérations du recensement. *Dictionnaire du recensement de 2001, Référence*, n° 92-378-XPF au catalogue de Statistique Canada, 2002.

Division de la géographie. *Fichier de conversion des codes postaux (FCCP), Guide de l'utilisateur*, Statistique Canada, 1995.

Division de la géographie. *Fichier de conversion des codes postaux (FCCP), Guide de référence*, n° 92F0153GIF au catalogue de Statistique Canada, 2003.

Division de la statistique du revenu. *Le revenu au Canada, 2002*, n° 75-202-X1F au catalogue de Statistique Canada, 2002.

Division des normes. *Classification géographique type CGT 1991, volume I, La classification*, n° 12-571 au catalogue de Statistique Canada, 1992.

Division des normes. *Classification géographique type CGT 2001, La classification*, n° 12-571-XPB au catalogue de Statistique Canada, 2002.

